Design for Switched Optical Datapaths in Computer Systems

M.Glick

IRC –TR-04-020

July, 2004

IEEE ICTON

Intel Research Cambridge

Design for switched optical datapaths in computer systems

Madeleine Glick
Intel Research Cambridge, 15 JJ Thomson Ave., Cambridge CB3 0FD, UK
Tel: (+441223) 763441, Fax: (+441223) 763456, e-mail:madeleine.glick@intel.com

ABSTRACT
Recognising advantages of photonics for scalability and bandwidth we aim to design computer systems using switched optical datapaths. For acceptable system performance we must address issues of cost, stability and latency. Starting from the system design perspective, we have identified integrated digital control of optical devices and multiwavelength coding as enabling technologies. We present an architecture avoiding optical buffering, bit level synchronization and regeneration, which has applications to storage, desk area networks and optical chip interconnects. We draw upon well-understood computer networking concepts, and use wavelength striping of the packets to achieving low latency and high capacity.
Keywords: Optical networks, optical interconnect, SOA semiconductor optical amplifier, encoding techniques, WDM

1. INTRODUCTION

There has been much research investigating techniques to enable photonic packet switching; much has been published on Optical Burst Switching, Optical Packet Switching, and suitable photonics components [1, 2, 3]. However, major hurdles remain before the realization of a complete commercially viable system. There are significant outstanding technological challenges in providing buffering, synchronization and regeneration within the optical domain. There is little doubt, however, that photonics can offer significant advantages in scalability and bandwidth. Improvements in the capabilities, cost and availability of photonic components and transmission systems delivering low cost components are well matched to data-communications applications. Also as microprocessor clock frequencies, increase optical interconnects could be a solution to bandwidth limitations arising from signal attenuation, electromagnetic interference and crosstalk. [4]. These developments lead us to the investigation of switched optical data paths with applications to local, system, storage area networks and optical chip interconnects.

For the data network, essential criteria become low cost, latency and solutions that have a route to integration. Photonic components should be stable and reliable, preferably in a modular form which can be treated as a "black box" by the system designer. In these applications, we can perform optimization of the optical device control using a combination of optics, electronics and software aiming for a higher degree of modularisation of photonic devices, lowering cost and improving stability.

2. OBJECTIVE AND METHODS

In order to achieve optical packet switching with currently available components with a reasonably simple, and low cost, implementation suitable for applications in computing, we propose a *wavelength striped semi-synchronous optical interconnect* [5]. Restricting the applications considered to localized optical interconnects, this architecture allows acceptably high utilization in the optical data-path without the need for optical buffering, limits problems due to dispersion, non-linearities, and simplifies the processing of data in the network; we assume that issues concerning security, trust, classification, QoS, etc, which require significant computing power, are placed at the network boundaries and remain electronic.

We are building a testbed based on this architecture. This enables novel protocols and coding techniques across the multiple wavelengths to enhance burst mode clock recovery and optical balancing to be investigated to improve bandwidth efficiency and latency. The testbed will also be used to examine digital electronic control of the optical switch fabric in a system environment.

In the testbed a number of nodes are linked by means of multiple wavelength channels through an optical switching fabric as shown in the schematic diagram Fig. 1. The exact form of the switch fabric (ring, crossbar) only needs to be known to the central arbitration unit relating the control channel to the switch control. This control channel, on a separate wavelength which can be at a lower data rate, is used as the request / grant channel to implement the media access control (MAC). All routing information is transmitted at the control wavelength to avoid the necessity for complex high speed header processing.

The work reported in this paper was supported in part by the UK DTI-OSDA SOAPS project.
.

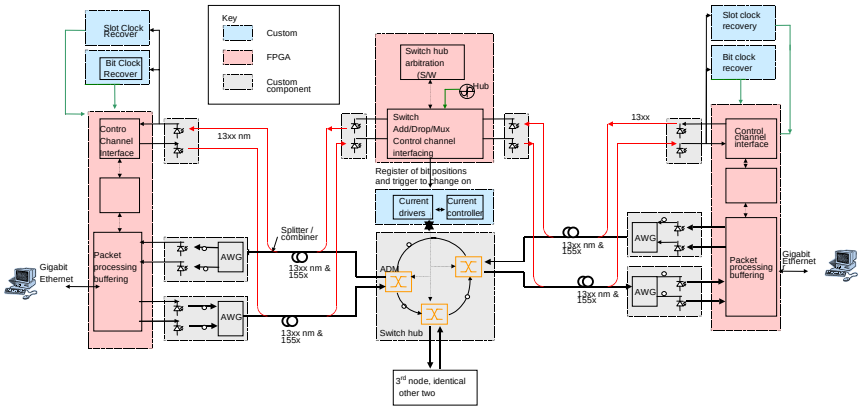

Figure 1. Schematic of the network showing wavelength striping and control channel

3. WAVELENGTH STRIPING

In a wavelength striped format (also called 'optical bus' [6] or 'bit parallel WDM' [7]) capacity is increased by simultaneously launching parallel pulses on different wavelengths (e.g. a bit from each wavelength channel is launched at the same time). This technique is particularly suitable for computer interconnects and high speed parallel systems where the requirement is for high data rates to a single end user. Wavelength striping could also offer advantages for novel coding techniques and potentially avoid serial to parallel, parallel to serial conversion.

Previous wavelength striped research has focused on networks with moderately long fibre lengths and used bit error rate (BER) as the impairment criteria, whereas we focus on the local area network and effects causing timing errors. Whereas most previous work has sought to compensate for or eliminate the delays our goal is to determine the maximum delay encountered by a bit during transmission and optimize for network efficiency as much as possible by choice of components and packet format.

4. SEMI SYNCHRONOUS SLOT FORMAT

The original LAN technologies all held packets in end-systems until successful transmission. They could be categorized as either synchronous or asynchronous. In the synchronous ones, nodes in the network received a continuous signal from the communications media to which they synchronized their transmissions at the bit level, and in which was encoded the information needed to implement Media Access Control (MAC) based on contention avoidance. In the asynchronous ones, such as Ethernet, a receiver would re-synchronize to the bit clock on a per packet basis and the MAC would implement contention resolution.

Un-buffered networks are found in the wireless realm. While local wireless networks often use similar asynchronous access techniques to local wired networks, experience in the wide area led to what can be categorized as semi-synchronous networks – that is synchronized at a time slot level, but asynchronous at the bit level. In these networks the variability in delay (jitter) introduced in radio propagation and/or mobility mean that is not possible to ensure that two packets from different sources arrive at the base-station (or satellite) with the required sub-bit timing accuracy to run synchronously.

Drawing on this previous work, we consider a semi-synchronous optical network composed of point-to-point links and a central switch.

5. CONTROL CHANNEL

The control channel is on a separate wavelength (1300 nm in our implementation). Thus, we avoid the need for o-e-o conversion or all optical processing of header and/or control information. It handles media access and timing. The channel is synchronized with fixed time slots for packet transmission; within the slot a suitably chosen guard time avoids bit loss, while the receiver node recovers the data at the bit level. The switch communicates slot timing and phase information to each of the network nodes so that the transmitters can lock onto the switch slot structure. It is also used as the request / grant channel to implement the MAC. As with wireless communications we can consider MAC layers implementing reservation and contention mechanisms.

6. BIT DELAY, GUARD BAND AND NETWORK EFFICIENCY

In the wavelength striped format a limiting factor is bit skew due to group velocity dispersion in the fibre. We also must account for delays due to environmental effects such as change in temperature. We therefore, either introduce guard bands, gaps between packets or sacrificial packet preambles. Presuming bit asynchronous data being transported in a slot sized to take a standard Ethernet frame (12000 bits), to achieve 90% utilisation, bit loss due to slip must be less than 1200 bits. At data rates of 10Gbps this would be equivalent to 120ns of bit slip, and at 100Gbps 12ns slip. If the guard band can be sufficiently short to allow adequate network efficiency, bit level synchronizers are not needed at the switch, considerably reducing the complexity of the network. In this way, timing constraints are not as severe although fast clock recovery is required at the receiving node (discussed below). We have estimated that the worse case delay for our data rates and link is well within our 1200 bit limit [8].

7. OPTICAL SWITCH FABRIC

To provide the switching function, a special add-drop device, which has the same functionality as a crosspoint switch, has been designed and constructed using semiconductor optical amplifier (SOA) gates as shown in Fig. 2 [9]. These gates are attractive for switching applications because of their high on-off ratios, wideband operation and ability to switch on fast timescales. Details of the switch are described in detail elsewhere [9, 10]

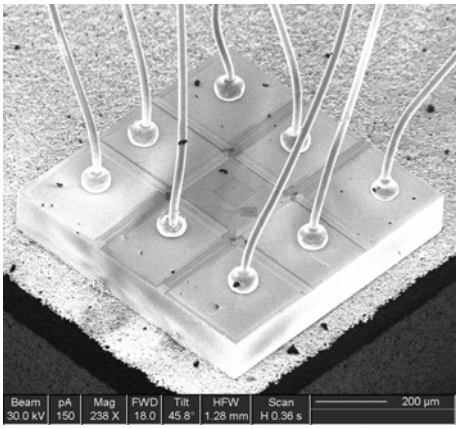

Figure 2a. Secondary ion image of a two input two output optical switch with total internal reflecting element mirrors formed by focussed ion beam etching.

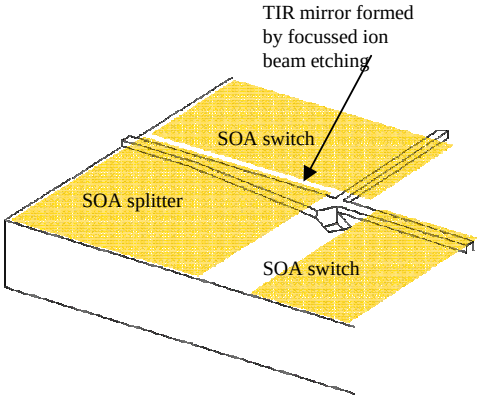

Figure 2b. Schematic diagram of the basic 1x2 as implemented four times to form the 2x2 switch matrix.

The add-drop device performance has been assessed using WDM sources. Switching between the add-drop modes can be achieved with a 4.5ns transition from 0-30mA on the appropriate gate. A device power consumption of 170mW is recorded for all valid operating states. Operation with less than 5 dB loss is shown from 1535nm to 1560nm as in Fig. 3 [11].

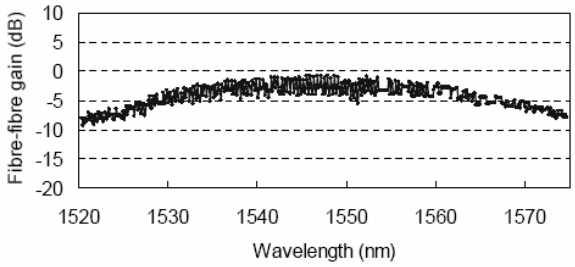

Figure 3. Wavelength dependant gain measured through a crosspoint add-drop gate

The good optical spectrum from 1535nm to 1560nm and operation of 4 simultaneous wavelengths shows that WDM operation with at least 4 wavelengths is achievable at a data rate of 2.5Gbit/s. A full optical interconnect can therefore be envisaged using wavelength striping for enhanced overall network efficiency and flexibility,

using low cost components

In recent results we have shown [12] wavelength striped packets at 80Gb/s aggregate bit rate routed with a separate 100Mb/s control channel. In this experiment an FPGA drove the broadband Mach Zehnder and SOA switches.

8. MULTIWAVELENGTH CODING

An important scaling issue is how multi-wavelengths can be effectively used to aggregate bandwidth. This has not been thoroughly investigated for packet switching due to previous emphasis on WDM for long haul applications. In an optically switched network packets received consecutively may originate from different source nodes at different distances. Traditional optical serial networks use phase locked loops (PLLs) to recover the clock from a serial transmission that has been coded to assist PLL locking. PLLs require a considerable preamble to ensure a lock when the sources are not synchronous. The insertion of a significant preamble reduces the efficiency of the network. Several research groups have recognised this problem [13, 14, 15]. We are investigating asynchronous techniques to minimize or eliminate preamble requirements thereby maximizing efficiency. We intend to exploit the ability to use multiple wavelengths for high capacity by looking into delay insensitive n-of-m codes [16] for applications using striped wavelengths. n-of-m codes combine significant level of robustness as well as allowing for bit and symbol clock recovery without preamble.

In addition, a drawback to SOA use in multiwavelength systems is that their carrier lifetimes, in the nsec regime, lead to transient gain variation, crosstalk problems due to cross gain modulation, at Gbps data rates. Several ways to address this have been studied [17]. Multiwavelength coding techniques will also be investigated to address this problem. Initial results from sending asynchronous data on multiple wavelengths through the SOA indicate that optical power balancing by proper coding could enhance system performance [18].

9. CONCLUSIONS

Acceptable levels of utilization should be obtained in a wavelength striped semi-synchronous network using high speed SOA based optical switches, making this solution a good candidate for local area, storage area networks and optical interconnects. The use of a single lower speed WDM channel (in our case at 1300 nm) to carry the control functions for the network is a key feature of the scheme. High speed channels are used to carry the data, allowing the most complex electronic signal processing to use slower speed and lower cost circuitry to control the network thus achieving reduced component count and power consumption. Advances in low cost, low footprint uncooled optical transceiver technology, currently being developed for point to point synchronous data transmission and digital electronic control can be expected to make this solution more economically attractive. In addition, novel use of the multiple wavelengths could more efficiently solve problems with high speed bit level clock recovery.

Efficient and viable systems using multiwavelength photonic packet switching will be enabled by appropriate management of the photonic, electronic and software interfaces.

ACKNOWLEDGEMENTS

The work described above is the product of the Optical Systems work at Intel Research Cambridge in collaboration Derek McAuley, and with Tao Lin, Guy Roberts, Laura James, Kevin Williams, Richard Penty and Ian White of the Photonics System Group, Engineering Department, University of Cambridge UK and Andrew Moore and Simon Moore of the Computer Lab, University of Cambridge.

REFERENCES

[1] M. Hamdi, et al. Eds.: *IEEE Jour. on Selected Areas in Communication, High-performance optical switches/routers for high-speed internet, vol. 21, pp.1013-12 , Sept. 2003.*

[2] C. Guillemot et al.: Transparent optical packet switching: The European ACTS KEOPS project approach, *J. Lightwave Technol., vol. 16, pp. 2117-2134, 1998.*

[3] I. M. White et al.: A Summary of the HORNET Project: A Next-Generation Metropolitan Area Network, *JSAC, vol. 21 pp.1478-1494, Nov. 2003*.

[4] D. A. B. Miller: Rationale and Challenges for Optical Interconnects to Electronic Chips, *Proc. IEEE, vol. 88, pp. 728-749, 2000.*

[5] Derek McAuley: An Optical LAN, *in Computer Systems: Theory, Technology and Applications*, Herbert, A., Spärck-Jones, K eds..; Springer Verlag, 2003.

[6] M. R. Loeb and G. R. Stilwell: High speed data transmission on an optical fibre using a byte wide WDM system, *IEEE Jour. Lightwave Technology, vol. 6, p. 1306, 1988.*

[7] L. Bergman, C. Yeh, and J. Morookian: Advances in Multichannel MultiGbytes/s Bit-Parallel WDM Single Fibre Link, *IEEE Transactions on Advanced Packaging, vol. 24, p. 456, 2001.*

[8] L.B. James, G.F. Roberts, M. Glick, D. McAuley, K.A.Williams, R.V. Penty, I.H. White: Wavelength Striped Semi-syncronous Optical Local Area Networks, *Proc. London Communications Symposium, Sept 8-9, 2003.*

[9] K A Williams, G F Roberts, R V Penty, I H White, M Glick, D McAuley, D J Kang and M G Blamire: Monolithic 2x2 Amplifying Add/Drop Optical Switch for Data Networking, *Proc. SIOE 2003, Cardiff, UK.*

[10] G.F. Roberts, K.A. Williams, R.V. Penty, I.H. White, M. Glick, D. McAuley, D.J. Kang, M. Blamire: Monolithic 2x2 Amplifying Add/Drop Switch for Optical Local Area Networking, *Proc. ECOC 03, Rimini, Italy, Sept. 2003.*

[11] I.H. White, R.V.Penty, J.Hankey, K.A.Williams, G.F. Roberts, M. Glick, D. McAuley: Optical local area networking using CWDM, *Proc. SPIE ITCom, San Diego, 2003 Sept.*

[12] T. Lin, W. Tang, K. A. Williams, G. F. Roberts, L. B. James, R. V. Penty, I. H, White, M. Glick, D. McAuley: 80Gb/s Optical Packet Routing for Data Networking using FPGA Based 100Mb/s Control Scheme, *submitted to ECOC 2004.*

[13] I. M. White: *A New Architecture and Technologies for High-Capacity Next-Generation Metropolitan Networks*, Thesis, Stanford University, 2002.

[14] G. Georgiou et al: Clock and Data Recovery IC for 40-Gb/s Fiber Optic Receiver, *IEEE Jour. SSC, vol. 37,*
pp. 1120 – 1125, Sept. 2002.

[15] C. K. Yang and M. Horowitz: A 0.8 micron CMOS 2.5 Gb/s Oversampling Receiver and Transmitter for Serial Links, *IEEE Jour. SSC, vol. 3 , pp. 2015 – 2023, Dec. 1996.*

[16] W.J. Bainbridge, W.B. Toms, D.A. Edwards, S.B. Furber: Delay-Insensitive, Point-to-Point Interconnect using m-of-n codes, *Proc. Async'2003, pp. 132-140, Vancouver, May 2003.*

[17] L. Spiekman: Semiconductor Optical Amplifiers, *in Optical Fiber Telecommunications IVa: Components*, I Kaminow and T. Li, Eds., Academic Press 2002.

[18] G.F. Roberts, K.A. Williams, R.V. Penty and I.H. White, M. Glick and D. McAuley, D.J. Kang and M.G. Blamire: Integrated SOA based add-drop multiplexer for WDM optical packet routing with reduced distortion penalty, *Proc. CLEO 04, San Fransisco, Ca,. May 2004.*